Iranian Journal of Animal Science Research
Vol. 12, No. 2, Summer 2020, p. 241-250

نشریه پژوهشهای علوم دامی ایران
جلد ۱۲، شماره ۲، تابستان ۱۳۹۹، ص. ۲۵۰-۲۴۱

مقاله علمی- پژوهشی

مقایسه روش‌های GBLUP و بیزی در برآورد ارزش‌های اصلاحی ژنومی با معماری‌های مختلف ژنتیکی

رضا سید شریفی [1*]، فاطمه علاء نوشهر [2]، نعمت هدایت ایوریق[1]، جمال سیف دواتی[1]

تاریخ دریافت: ۱۳۹۸/۰۳/۰۱
تاریخ پذیرش: ۱۳۹۸/۰۷/۰۷

چکیده

انتخاب ژنومی با ا ستفاده از ن شانگرهای با تراکم بالا، بخ صوص SNPهایی که کل ژنوم را پو شش می‌دهند و اغلب در عدم تعادل پیو ستگی با QTLهای مجاور خود قرار دارند، ارزش ژنتیکی کل را پیش‌بینی می‌کند. این پژوهش با هدف برر سی تأثیر عواملی چون تراکم نشانگرها، تعداد QTL، وراثت‌پذیری صفت و نوع توزیع اثر QTL بر صحت برآورد ارزش‌های ا صلاحی ژنومی با ا ستفاده از داده‌های شبیه سازی در گو سفند انجام گرفت. به همین منظور ژنومی متشکل از سه کروموزوم، هر یک به طول ۱۰۰ سانتی‌مورگان با سه مقدار وراثت‌پذیری برای صفت مورد بررسی (۰/۱، ۰/۳ و ۰/۵) و سه پنل ن شانگری (۵۰۰، ۱۰۰۰ و ۱۵۰۰) در سه سطح تعداد QTL (۵۰، ۱۰۰ و ۱۵۰) با دو اثر توزیع یکنواخت و گاما برای QTL شبیه سازی شدند. صحت ارزش‌های ا صلاحی ژنومی برآورد شده با ا ستفاده از پنج روش GBLUP، بیزA، بیزB، بیزC و بیز LASSO مورد مقایسه قرار گرفتند. نتایج این تحقیق نشان داد که هر چه تراکم نشانگر و وراثت‌پذیری صفت افزایش یافته و تعداد QTL مؤثر بر صفت کمتر باشد، صحت ارزش اصلاحی برآورد شده بالاتر خواهد بود. در بین روش‌های آماری زمانی که تعداد QTL مؤثر بر صفت پایین است و توزیع اثر QTL گاما در نظر گرفته شد، پیش‌بینی ارزش‌های اصلاحی ژنومی با روش بیزB عملکرد بهتری داشت.

واژه‌های کلیدی: GBLUP، بیزA، بیزB، بیزC، بیزLASSO، ارزش اصلاحی ژنومی.

مقدمه

در روش‌های کلاسیک اصلاح دام، انتخاب برای صفات مهم اقتصادی با استفاده از اطلاعات شجره‌ای به همراه رکوردهای فنوتیپی خود فرد، آیندگان (فرزندان و نوه‌ها) و گذشتگان (والدین و سایر افراد خویشاوند نسل‌های قبل) انجام گرفته و بهترین پیش‌بینی ناااریب خطی از ارزش‌های اصلاحی (BLUP)[3] فرد حاصل می‌شود (۹). با پیشرفت ژنتیک مولکولی استفاده از اطلاعات نشانگرهای DNA در برنامه‌های اصلاح‌نژاد، تحت عنوان انتخاب به کمک نشانگر (MAS)[4] پیشنهاد شد (۳). بعد از آن شالوده انتخاب بر اساس کل ژنوم در مقاله‌ای که توسط موویسن و همکاران (۱۰) انتشار یافت، ارائه شد. در این روش می‌توان ارزش‌های اصلاحی ژنومی تمام افراد را با استفاده از یک مدل خطی به‌صورت تابعیت فنوتیپ‌ها از تمام نشانگرهای متراکم که کل ژنوم را پوشش می‌دهند با صحت بالا پیش‌بینی کرد. مهم‌ترین مرحله این روش برآورد اثرات آللی می‌باشد، زیرا هر چه اثرات QTL [5] به نحو صحیح‌تری برآورد شوند، ارزش‌های اصلاحی پیش‌بینی شده بیشتر معرف ظرفیت ژنتیکی حیوان بوده و رتبه‌بندی حیوانات به‌طور صحیح‌تری انجام خواهد شد و در نتیجه پیشرفت ژنتیکی و پاسخ به انتخاب بیشتری مورد انتظار خواهد بود.

عوامل مختلفی می‌توانند صـحت ارزش‌های اصـلاحی ژنومی را تحت تأثیر قرار دهند. این عوامل شـامل مدل آماری مورد اسـتفاده، توزیع اثرات QTL های مؤثر بر صـفت، مقدار عدم تعادل پیوسـتگی، تراکم نشـانگرها، وراثت‌پذیری صـفت، تعداد داده‌های فنوتیپی در جمعیت مرجع (گروهی که دارای اطلاعات فنوتیپی بوده و از اطلاعات آنها جهت برآورد اثرات نشانگری استفاده می‌شود)، فاصله زمانی (تعداد

۱-دانشیار گروه علوم دامی دانشگاه محقق اردبیلی، اردبیل، ایران.
۲- دانش آموخته دکتری ژنتیک و اصلاح دام، دانشگاه تبریز، تبریز، ایران.
(*-ایمیل نویسنده مسئول: reza_seyedsharifi@yahoo.com)
DOI:10.22067/ijasr.v12i2.80862

3 - Best Linear Unbiased Prediction
4 - Marker Assisted Selection
5 - Quantitative Trait Loci

نســـل) بین جمعیت مرجع و جمعیت تأیید (گروهی که فاقد اطلاعات فنوتیپی بوده و ارزش اصـــلاحی ژنومی آنها تنها بر اســـاس اطلاعات ژنوتیپی برآورد می‌شود) می‌باشند (۱۸).

تحقیقات نشان داده است که با افزایش تراکم نشانگر در سطح ژنوم، برآوردهایی با صحت بالاتر حاصل خواهد شد. اصولاً افزایش تعداد نشانگرها در سطح ژنوم احتمال پیوستگی بین نشانگر و QTL را افزایش می‌دهد که این مسئله باعث افزایش صحت برآوردها می‌شود (۲). همچنین در وراثت‌پذیری بالاتر، ارزش‌های اصـلاحی افراد سـهم بیشـــتری در بروز رکوردهای فنوتیپ داشـــته و لذا در برآورد اثرات نشـانگرها با اسـتفاده از اطلاعات فنوتیپی صـحت بیشـتری مشـاهده خواهد شد (۱۰).

امروزه تلاش پژوهشگران بر آن است تا با شناسایی عوامل مؤثر بر صحت برآوردهای ارزش‌های اصلاحی ژنومی بهترین بهره‌برداری از داده‌های موجود انجام شـــود. از آنجا که توالی‌یابی کل ژنوم و در اختیار داشتن ارزش ژنتیکی افراد به خصوص در گوسفند به دلیل نوع سـیسـتم پرورش، بسـیار سـخت و هزینه‌بر اسـت، انجام مطالعات با استفاده از شبیه‌سازی روشی منطقی برای چنین موضوعی خواهد بود. لذا پژوهش حاضر با هدف بررسی صحت برآورد ارزش‌های اصلاحی ژنومی صفات پیوسته در معماری‌های مختلف ژنتیکی شامل توزیع‌های آماری مختلف اثرات ژنی، تعداد متفاوت QTL، ســطوح متفاوت وراثت‌پذیری و تراکم مختلف نشـــانگری با اســـتفاده از روش‌های GBLUP و بیزی شــــامل بیزA، بیزB، بیز C و بیز LASSO انجام گرفت. تا علاوه بر مقایســـه نحوه عملکرد روش‌های مختلف در معماری‌های مختلف ژنتیکی، بتوان میزان تراکم نشانگر و تعداد QTL کارآمدی را برای برنامه‌های شبیه‌سازی جمعیت‌های گوسفند معرفی کرد.

مواد و روش‌ها

شبیه‌سازی جمعیت

جهت ایجاد عدم تعادل پیوســـتگی مورد نظر بین نشـــانگرها و QTLها از روش ویلامسن و همکاران (۱۸) استفاده شد. در این روش از عامل کوچک بودن جمعیت و رانش، جهت ایجاد عدم تعادل پیوستگی استفاده می‌شود. به همین منظور جمعیت پایه کوچکی (G_0) با ۱۰۰ رأس دام شامل ۵۰ رأس نر و ۵۰ رأس ماده در نظر گرفته شد. فراوانی آللی اولیه برای چندشکلی‌های تک نوکلئوتیدی در نسل پایه، ۰/۵ لحاظ گردید. گامت‌های والدی بر اســـاس فرض عدم تعادل پیوستگی و با استفاده از روش تابع مکان‌یابی هالدان (۸) شبیه‌سازی شدند و سپس گامت‌های نوترکیب ایجاد شده به طور تصادفی انتخاب و آمیخته شـــدند تا فرد جدیدی از نســـل G_1 ایجاد شـــود. روش مورد اســـتفاده تا ایجاد نســـل ۵۰ام پیگیری شـــد. بعد از ۵۰ نســـل آمیزش تصـــادفی اندازه جمعیت به تعداد ۱۰۰۰ فرد (۵۰۰ نر و ۵۰۰ ماده) در نســـل ۵۱ گســـترش یافت. این اندازه جمعیت برای ۹ نســـل بعدی (تا نسل ۶۰) ثابت ماند. تنها افراد نسل ۵۱ دارای رکورد فنوتیپی بودند و گروه مرجع (گروهی که از اطلاعات آنها جهت برآورد اثرات نشانگری استفاده می شود) را تشکیل می‌دادند و افراد سایر نسل‌ها (۵۲ تا ۶۰) فاقد اطلاعات فنوتیپی بودند و گروه تأیید (گروهی که فاقد رکورد فنوتیپی بوده و ارزش اصـــلاحی ژنومی آنها تنها بر اســـاس اطلاعات ژنوتیپی برآورد می‌شـــود) را به وجود آوردند. شـــبیه‌ســـازی داده‌ها با استفاده از نرم‌افزار qmsim (۱۵) انجام شد.

جهت سنجش صحت شبیه‌سازی از روش بررسی روند عدم تعادل پیوستگی در هر نسل استفاده شد. برای این منظور مقدار مورد انتظار عدم تعادل پیوستگی در هر نسل بر اساس فرمول اسوید (۱۷) محاسبه شد: رابطه ۱

$$E(r_n^2) = \frac{1}{1+4N_e c\left[\frac{(1-c/2)}{(1-c)^2}\right]}\left\{1-\left[(1-\frac{1}{2N_e})(1-c)^2\right]^n\right\}$$

در رابطه فوق $E(r_n^2)$ میزان مورد انتظار عدم تعادل پیوســـتگی در نسل nام، N_e اندازه مؤثر جمعیت و c میزان نرخ نوترکیبی بر اساس مورگان هستند.

شبیه سازی ژنوم

در این مطالعه ژنومی شامل سه کروموزوم هر کدام به طول یک مورگان در نظر گرفته شـــد. روی هر کروموزوم ســـه ســـطح تراکم چندشـکلی تک نوکلئوتیدی دو آللی شـامل (۵۰۰، ۱۰۰۰ و ۱۵۰۰) در مورگان با مقادیر مختلف qtl شامل ۵۰ (پایین)، ۱۰۰ (متوسط) و ۱۵۰ (بالا) qtl با فواصل یکسان نشانگری گسترده شدند. همچنین جهت مطالعه تأثیر وراثت‌پذیری و نحوه توزیع واریانس qtl بر صحت برآورد ارزش‌های اصلاحی ژنومی از سه صفت گوسفند با وراثت‌پذیری پایین (۰/۱)، متوسط (۰/۳) و بالا (۰/۵) و دو نوع توزیع واریانس qtl شامل توزیع یکنواخت و گاما (β=۰/۴ و α=۱/۶۶) (۷) در شبیه‌سازی استفاده شدند. هم نشانگرها و هم qtlها دو آللی بودند.

بر این اساس ۱۸ سناریوی ترکیبی تعریف شد که شامل ۳ سطح برای تعداد qtl، ۲ ســـطح برای نحوه توزیع واریانس qtl و ۳ ســـطح برای تراکم نشانگرهای snp می‌باشد. همچنین سطوح وراثت‌پذیری در هر کدام از ســناریوها تغییر داده شــد تا تأثیر آنها بر صــحت برآورد ارزش‌های اصلاحی ژنومی بررسی شود.

پیش‌بینی ارزش های اصلاحی ژنومی

روش GBLUP

روش GBLUP به کار رفته در این مطالعه بر اســـاس یک مدل مختلط ساده طراحی شد و فرض گردید که در آن کلیه چندشکلی‌های تک نوکلئوتیدی بر صـــفت مورد مطالعه مؤثر می‌باشـــند. برای به کار بردن این روش از نرم‌افزار ASReml (۶) و مدل زیر استفاده شد:

$$y = \mu 1_n + Zu + e \qquad \text{رابطه ۲}$$

جدول ۱- ساختار جمعیتی و متغیرهای استفاده شده در شبیه‌سازی

Table 1- Population structure and simulation parameters

متغیر Parameter	ارزش Value
اندازه مؤثر جمعیت Effective population size (Ne)	100
طول ژنوم Genome length	300
تعداد چندشکلی‌های تک نوکلئوتیدی Number of SNP markers	500, 1000, 1500
تعداد QTL Numbers of QTL	50, 100, 150
توزیع اثر QTL QTL effects distributions	یکنواخت و گاما (β=۰/۴ و α=۱/۶۶)
وراثت پذیری صفت Heritability	0.1, 0.3, 0.5
جمعیت مرجع Training set	همه افراد نسل ۵۱
جمعیت تأیید Validation set	همه افراد نسل ۵۲ تا ۶۰
تعداد نسل در جمعیت عدم تعادل پیوستگی Number of generations in linkage disequilibrium population	50
اندازه جمعیت عدم تعادل پیوستگی در هر نسل Population size of linkage disequilibrium per generation	۱۰۰ رأس دام (۵۰ رأس نر و ۵۰ رأس ماده)
اندازه جمعیت Population size	۱۰۰۰ رأس دام (۵۰۰ رأس نر و ۵۰۰ رأس ماده)
نوترکیبی Recombination	تابع مکان‌یابی هالدان
تعداد نسل جمعیت Number of population generation	60
تعداد تکرار Number of repetitions	10

که در آن y بردار ارزش های فنوتیپی، $\boldsymbol{\mu}$ میانگین فنوتیپی، n تعداد رکورد ها، $\mathbf{1}_n$ بردار یکا به ابعاد n×1، $\boldsymbol{u}$ بردار در برگیرنده ارزش های اصلاحی ژنومی، $\boldsymbol{Z}$ ماتریس ضرایب ارتباط دهنده مشاهدات به بردار ارزش‌های اصلاحی و e بردار اثرات تصادفی باقیمانده می‌باشد. اثرات ژنتیک افزایشی چندشکلی‌های تک نوکلئوتیدی (g) دارای توزیع نرمال $\boldsymbol{N(0,\sigma_g)}$ فرض شد که Z ماتریس روابط خویشاوندی برای همه چندشکلی‌های تک نوکلئوتیدی می‌باشد که $\boldsymbol{\sigma_g}$ انحراف معیار اثرات ژنتیکی SNPها می‌باشد. باید توجه داشت که Z بر اساس احتمالات IBS بین هر دو دام محاسبه شده است و برای تمام دام‌های دارای ژنوتیپ (دام‌های جمعیت مرجع و جمعیت تأیید) تعمیم یافته است. برای محاسبه مجموع روابط آللی بین هر دو دام از روش نجاتی جوارمی و همکاران (۱۲) استفاده شد. برای پیش‌بینی اثرات افزایشی چندشکلی‌های تک نوکلئوتیدی از مدل‌های مختلط زیر استفاده شد:

$$\begin{bmatrix} 1_n' 1_n & 1_n' Z \\ Z' 1_n & Z'Z + I\alpha \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' y \\ Z' y \end{bmatrix} \qquad \text{رابطه ۳}$$

که در آن $\boldsymbol{\alpha} = \frac{\sigma_e^2}{\sigma_g^2}$ و I ماتریس همانی می‌باشد.

روش بیز A

در همه روش‌های بیزی از مدل آماری زیر استفاده شد:

رابطه ۴ $y = \mu 1_n + \sum_{i=1}^{m} X_i g_i + e$

که در آن y بردار ارزش های فنوتیپی، μ میانگین فنوتیپی، n تعداد رکوردها، $\mathbf{1_n}$ بردار یکا به ابعاد 1×n، $\boldsymbol{X_i}$ ماتریس ژنوتیپ‌ها برای i امین چندشکلی تک نوکلئوتیدی، g_i اثر جایگزینی آللی i امین چندشکلی تک نوکلئوتیدی و e بردار اثرات تصادفی باقیمانده می‌باشد.

در روش BayesA همانند روش GBLUP فرض می‌شود که تمامی SNPها دارای اثر، حتی جزئی بوده و تعدادی از SNPها در عدم تعادل پیوستگی با QTLهایی با اثرات متوسط تا بزرگ قرار دارند. اثرات SNPها از توزیع نرمال با واریانس جداگانه برای هر SNP از توزیع کای دو معکوس، نمونه‌گیری می‌شود (۱۰). برای به کار بردن این روش از معادلات زیر استفاده شد:

رابطه ۵ $f(\sigma^2_{g_j} | v, s^2) \sim \chi^{-2}(v, s^2)$

$v = 4.2,\ \hat{\sigma}^2_{SNP} = \dfrac{\hat{\sigma}^2_a}{(1-\pi)\sum_{m=1}^{M} 2p_m q_m}$

$s^2 = \dfrac{\hat{\sigma}^2_{SNP}(v-2)}{v}$

که در آن v دلالت بر درجه آزادی و s مربوط به پارامتر معیار است.

روش بیز B

در این روش اکثر SNP های موجود در منطقه ژنومی QTLای نداشته و بنابراین اثرات آنها برابر با صفر می‌باشد، در حالی‌که تعداد اندکی از آنها (π−۱) در عدم تعادل پیوستگی با QTL قرار داشته و دارای اثر می‌با شند. در نتیجه اثرات غیر صفر در عدم تعادل پیو ستگی بالایی با QTL قرار دارند. بیز b تحت معادلات زیر اجرا شد:

رابطه ۶

$$f(\sigma^2_{g_j} | v, s^2, \pi) \begin{cases} \sim \chi^{-2}(v, s^2) & with\ probability\ (1-\pi) \\ 0 & with\ probability\ \pi \end{cases}$$

روش بیز C

در روش بیز C فرض بر این است که تنها بخشی از چندشکلی‌های تک نوکلئوتیدی (π) بر صفت اثر دارند. در واقع بیز C شکلی از بیز B است، فقط به جای استفاده از توزیع پیشین t برای تأثیرات نشانگری، از توزیع نرمال استفاده می‌کند. برای به کار بردن این روش از مدل زیر استفاده گردید: رابطه ۷

$\boldsymbol{y = \mu 1_n + \sum_{i=1}^{m} x_i g_i + e}$

که در آن y بردار ارزش‌های فنوتیپی، μ میانگین فنوتیپی، n تعداد رکوردها، 1_n بردار یکا به ابعاد 1*n، x_i بردار ژنوتیپ‌ها برای iامین چندشکلی تک نوکلئوتیدی، g_i اثر جایگزینی آللی iامین چندشکلی تک نوکلئوتیدی و e بردار اثرات باقیمانده است که دارای توزیع نرمال می‌باشد.

روش بیز LASSO

در روش بیز Lasso از تابع نمائی دوگانه برای توزیع اثرات QTL و SNP تحت مدل بیزی استفاده شد:

رابطه ۸ $DE(b_j | \lambda, \sigma^2_e) = \int N(b_j | 0, \sigma^2_e, \tau^2_j)$

$Exp\left(\tau^2_j \middle| \frac{\lambda^2}{2}\right) \delta\sigma^2_{bj}$

که در معادله فوق λ ناشناخته فرض شده و به آن توزیع گاما با پارامتر شکل λ و مقیاس τ اختصاص داده می‌شود.

برای محا سبه ارزش ا صلاحی واقعی حیوانات از رابطه ۱۹ ستفاده شد:

رابطه ۹ $TBV_i = \sum_{j=1}^{n} \boldsymbol{x_{ij} b_j}$

که در معادله فوق TBV_i ارزش اصلاحی واقعی افراد، n تعداد QTLهای مؤثر بر صفت، $\boldsymbol{x_{ij}}$ تعداد آلل‌های ۱ که فرد i در لوکوس j حمل می‌کند و $\boldsymbol{b_j}$ اثر jامین QTL می‌باشد. در این مطالعه همه ۱۰۰۰ رأس دام نسل ۵۱ به عنوان جمعیت مرجع به کار گرفته شدند و از فنوتیپ دام‌ها در این نسل برای برآورد ارزش‌های ا صلاحی ژنومی در جمعیت تأیید استفاده شد. ارزش اصلاحی ژنومی برآورد شده (ارزش اصلاحی ژنومی افراد گروه تأیید) از مجموع آثار نشانگرها با توجه به ژنوتیپ آنها به صورت زیر محاسبه شد:

رابطه ۱۰ $GEBV = X.\hat{g}$

که در آن X ماتریس طرح می‌باشد که به ژنوتیپ دام‌ها اشاره دارد و $\hat{\boldsymbol{g}}$ بردار اثرات برآورد شده برای چندشکلی‌های تک نوکلئوتیدی می‌باشد.

صحت برآورد ارزش‌های اصلاحی ژنومی از همبستگی بین ارزش‌های ا صلاحی واقعی و ارزش‌های ا صلاحی ژنومی برآورد شده (tbv,gebv) محاسبه شد. هر آزمایش ۱۰ بار تکرار و سپس میانگین و خطای استاندارد برآوردها گزارش شد.

در اجرای روش های بیزی از روش نمو نه‌برداری گیبس و نرم افزار R ورژن ۱.۱-۳ (۱۳) ا ستفاده شد. برای هر تحلیل یک زنجیره مارکوف مونت کارلو تا رسیدن به همگرایی به طول ۲۱۰ هزار چرخه شامل ۱۰ هزار چرخه اولیه جهت گرم شدن و ۲۰۰ هزار چرخه ا صلی به کار رفت که در هر ۵ دور یک بار نتایج ذخیره شد و در پایان نتایج، ۴۰ هزار چرخه ذخیره گردید.

نتایج و بحث

میانگین عدم تعادل پیو ستگی (r^2) افراد ن سل آخر جمعیت (ن سل ۵۰ ام) ۰/۰۱۱ ± ۰/۱۹۱ بدست آمد. این در حالی است که مقدار مورد انتظار آن توسط فرمول اسوید (۱۶) ۰/۲۱۹ برآورد گردید. این مقدار نشان می‌دهد که به طور میانگین بیش از ۸۷ درصد از حداکثر مورد

انتظار، محقق گردیده اســـت که میزان مطلوبی اســـت. میانگین عدم تعادل پیوستگی با مطالعات ویلیامسن و همکاران (۱۷) مطابقت دارد.

تأثیر تعداد QTL بر صحت برآورد ارزش‌های اصلاحی ژنومی

صحت ارزش‌های اصلاحی ژنومی برآورد شده با سه سطح تعداد QTL و وراثت‌پذیری در جدول ۲ ارائه شـــده اســـت. در هر پنج روش آماری با افزایش تعداد qtlها از ۵۰ به ۱۵۰، صـحت ارزش‌های ژنومی کاهش یافت. در هر ســـه ســـطح وراثت‌پذیری با تعداد qtl پایین، بالاترین صـحت برآورد مربوط به روش بیز b و کمترین مقدار توسـط روش gblup حاصل شد. این نتایج با نتایج دیتویلر و همکاران (۵) که گزارش کرده‌اند با افزایش تعداد qtl مؤثر بر صفت، صحت برآوردهای ارزش‌های اصلاحی ژنومی کاهش می‌یابد، مطابقت دارد.

دلیل این امر را می‌توان به توزیع مقدار محدودی واریانس ژنتیکی بر تعداد زیادی qtl دانست که در نتیجه سهم هر qtl در ارزش ژنتیکی کل کاهش یافته اســـت و قدرت مدل‌ها در برآورد تأثیرات کم خواهد شد (۵).

جدول ۲– صحت ارزش‌های اصلاحی ژنومی برآورد شده با سه سطح تعداد QTL و وراثت‌پذیری

Table 2– The estimated genomic accuracy (se) for three QTL numbers and heritability in same SNP density

وراثت پذیری Heritability	تعداد QTL N_{QTL}	مدل آماری Statistical model GBLUP	BayesA	BayesB	BayesC	Bayes LASSO
0.1	50	0.679(0.01)	0.685(0.01)	0.757(0.01)	0.756(0.02)	0.757(0.02)
	100	0.673(0.01)	0.681(0.01)	0.743(0.01)	0.741(0.02)	0.740(0.02)
	150	0.669(0.01)	0.672(0.01)	0.658(0.01)	0.652(0.02)	0.652(0.02)
0.3	50	0.689(0.01)	0.695(0.01)	0.776(0.01)	0.770(0.02)	0.767(0.02)
	100	0.681(0.01)	0.687(0.01)	0.768(0.01)	0.764(0.02)	0.762(0.02)
	150	0.678(0.01)	0.675(0.01)	0.657(0.01)	0.652(0.02)	0.650(0.02)
0.5	50	0.756(0.01)	0.759(0.01)	0.840(0.01)	0.838(0.01)	0.836(0.01)
	100	0.727(0.01)	0.734(0.01)	0.821(0.01)	0.818(0.01)	0.817(0.01)
	150	0.715(0.01)	0.727(0.01)	0.743(0.01)	0.739(0.01)	0.735(0.01)

مطالعات در این زمینه نشـــان داده اســـت که وقتی تعداد qtlهای مؤثر بر صـــفت زیاد باشـــد، روش gblup مشـــابه و یا حتی بهتر از روش‌های بیزی عمل می‌کند (۵ و ۱۹). با وجود این، به دلیل اینکه روش‌های بیزی معرفی بهتری از نحوه توزیع اثرات QTL بر صـــفات، نسبت به روش GBLUP به صورت پیش‌فرض ارائه می‌دهد، در تعداد اندک qtl برتر عمل می‌کنند.

همچنین صـــحت برآورد ارزش‌های اصـــلاحی ژنومی صـــفات با وراثت‌پذیری بالا نسبت به وراثت‌پذیری پایین بیشتر است. دلیل این مسئله واضح است، زیرا هرچه وراثت‌پذیری صفت بیشتر باشد، فنوتیپ فرد به ارزش ژنتیکی فرد نزدیک‌تر بوده و در نتیجه آثار نشانگرها و به دنبال آن ارزش اصلاحی ژنومی افراد به طور صحیح‌تر برآورد می‌شود (۱۸).

تأثیر تراکم نشانگر بر صحت پیش‌بینی ارزش‌های اصلاحی ژنومی

صحت ارزش‌های اصلاحی ژنومی برآورد شده با سه سطح تعداد QTL و تراکم نشانگری در جدول ۳ ارائه شده است. با افزایش تراکم نشانگرها صحت پیش‌بینی توسط هر پنج روش آماری افزایش یافت. نتایج حاصـــل از این پژوهش با نتایج تحقیقات کالوس و همکاران (۲) و موبر (۱۱) مطابقت داشت. بیشترین صحت ارزش اصلاحی ژنومی توســـط روش بیز b در تراکم نشـــانگری ۱۵۰۰ و تعداد ۵۰ QTL به دست آمد.

جدول ۳- صحت ارزش‌های اصلاحی ژنومی برآورد شده با سه سطح تعداد qtl و تراکم نشانگری

Table 3–The estimated genomic accuracy (se) for three QTL numbers and SNP density Table

		مدل آماری Statistical model				
تعداد QTL N_{QTL}	تراکم نشانگری SNP density	GBLUP	BayesA	BayesB	BayesC	Bayes LASSO
50	500	0.685(0.01)	0.687(0.01)	0.769(0.01)	0.762(0.02)	0.760(0.02)
	1000	0.688(0.01)	0.693(0.01)	0.772(0.01)	0.768(0.02)	0.763(0.02)
	1500	0.694(0.01)	0.705(0.01)	0.788(0.01)	0.780(0.02)	0.779(0.02)
100	500	0.676(0.01)	0.677(0.01)	0.761(0.01)	0.758(0.02)	0.756(0.02)
	1000	0.679(0.01)	0.686(0.01)	0.765(0.01)	0.761(0.02)	0.759(0.02)
	1500	0.687(0.01)	0.699(0.01)	0.778(0.01)	0.773(0.02)	0.771(0.02)
150	500	0.669(0.01)	0.660(0.01)	0.649(0.01)	0.648(0.01)	0.641(0.01)
	1000	0.675(0.01)	0.673(0.01)	0.655(0.01)	0.651(0.01)	0.648(0.01)
	1500	0.692(0.01)	0.689(0.01)	0.668(0.01)	0.665(0.01)	0.661(0.01)

سولبرگ و همکاران (۱۶) گزارش کردند که با افزایش نشانگرها از ۱۰۰ به ۸۰۰ نشانگر در هر مورگان صحت ارزیابی ژنومی از ۶۹ درصد به ۸۶ درصد افزایش یافت. افزایش تعداد نشانگرها باعث افزایش میزان عدم تعادل پیوستگی بین ژن‌ها و نشانگرها می شود و بدین ترتیب صحت ارزیابی ژنومی را افزایش می‌دهد.

تأثیر وراثت‌پذیری بر صحت پیش‌بینی ارزش‌های اصلاحی ژنومی

صحت ارزش‌های اصلاحی ژنومی برآورد شده با سه سطح وراثت‌پذیری و تراکم نشانگری در جدول ۴ ارائه شده است. نتایج نشان داد که افزایش تراکم نشانگری در سطوح وراثت‌پذیری پایین (۰/۱) و بالا (۰/۵)، سبب افزایش صحت ارزش اصلاحی ژنومی می‌شود، اما در صفات با وراثت‌پذیری متوسط (۰/۳) افزایش تعداد نشانگرها بر صحت ارزیابی ژنومی تأثیری نداشت. بطوری‌که بیشترین صحت پیش‌بینی ارزش اصلاحی ژنومی در وراثت‌پذیری ۰/۵ با تراکم نشانگری ۱۵۰۰ توسط روش بیز B و کمترین صحت، توسط روش GBLUP در وراثت‌پذیری ۰/۱ با تراکم نشانگری ۵۰۰ به دست آمد. این نتایج با پژوهش‌های پیشین مطابقت داشت (۱۴ و ۷).

جدول ٤- صحت ارزش‌های اصلاحی ژنومی برآورد شده با سه سطح تراکم نشانگری و وراثت‌پذیری

Table4-The estimated genomic accuracy (se) for three SNP density and heritability

		مدل آماری Statistical model				
وراثت پذیری Heritability	تراکم نشانگری SNP density	GBLUP	BayesA	BayesB	BayesC	Bayes LASSO
0.1	500	0.673(0.01)	0.680(0.01)	0.742(0.01)	0.740(0.02)	0.739(0.02)
	1000	0.682(0.01)	0.693(0.01)	0.759(0.01)	0.751(0.02)	0.744(0.02)
	1500	0.695(0.01)	0.707(0.01)	0.763(0.01)	0.758(0.02)	0.754(0.02)
0.3	500	0.692(0.01)	0.695(0.01)	0.764(0.01)	0.760(0.02)	0.760(0.02)
	1000	0.693(0.01)	0.695(0.01)	0.765(0.01)	0.760(0.02)	0.760(0.02)
	1500	0.693(0.01)	0.696(0.01)	0.766(0.01)	0.761(0.02)	0.761(0.02)
0.5	500	0.715(0.01)	0.719(0.01)	0.779(0.01)	0.771(0.01)	0.771(0.01)
	1000	0.722(0.01)	0.729(0.01)	0.786(0.01)	0.781(0.01)	0.779(0.01)
	1500	0.738(0.01)	0.743(0.01)	0.791(0.01)	0.788(0.01)	0.785(0.01)

موویسن و همکاران (۱۰) و گودارد (۷) یکی از عوامل اصلی مؤثر بر صحت انتخاب ژنومی را وراثت‌پذیری صفت می‌دانند. بیشتر بودن وراثت‌پذیری، یعنی بیشتر بودن نسبت واریانس افزایشی به واریانس فنوتیپی، به معنای بیشتر بودن نقش ژن‌های با بیان افزایشی در ایجاد پراکنش در صفت است که باعث برآورد صحیح‌تر تأثیرات نشانگرها می شوند. صحت برآورد ارزش اصلاحی ژنومی صفات با وراثت‌پذیری

بالا نسبت به وراثت‌پذیری پایین بیشتر است. دلیل این مسئله واضح است زیرا هرچه وراثت‌پذیری صفت بیشتر باشد، فنوتیپ فرد به ارزش ژنتیکی فرد نزدیک‌تر بوده و در نتیجه آثار نشـــانگرها و به دنبال آن ارزش اصلاحی ژنومی افراد به طور صحیح‌تر برآورد می‌شود (۱۸ و ۷).

نتایج تأثیر واریانس QTL بر صـــحت برآورد ارزش‌های اصلاحی ژنومی

صـحت ارزش‌های اصـلاحی ژنومی در تراکم نشـانگری و تعداد QTL یکسان با توزیع واریانس یکنواخت و گاما در جدول ۵ ارائه شده اسـت. صـحت ارزش‌های اصـلاحی در حالت توزیع گاما تأثیر ژنی بهتری نسبت به توزیع یکنواخت داشت. همچنین، هنگامی که توزیع تأثیر ژنی گاما اسـت، روش‌های بیزی در مقایسـه با روش gblup از صـحت ارزش ژنومی بالاتری برخوردار هسـتند. در بین روش‌های بیزی، روش بیزb، بیزC و بیزLASSO در حالت توزیع گاما تأثیرات ژنی بهترین عملکرد را نشـان دادند اما تفاوت آنها از نظر آماری معنی‌دار نبود (p>۰/۰۵). نتایج حاصـل از این تحقیق با گزارش‌های موویسـن و همکاران (۱۰)، گودارد (۷)، سـولبرگ و همکاران (۱۶) و دیتویلر و همکاران (۵) مطابقت داشت.

دلیل این امر دو عامل می‌تواند باشـد: نخسـت آنکه توزیع پیش فرض روش‌های بیزی با توزیع گامای واریانس ژن‌های عمده اثر هماهنگی دارند که باعث برآورد صـحیح‌تر اثرات چندشـکلی‌های تک نوکلئوتیدی می‌گردد. همانطور که گودارد (۷) گزارش کرده است توزیع پیش فرض گاما در روش‌های بیزی برآوردهای صـحیح‌تری ایجاد می‌کند که با نتایج این مطالعه همخوانی دارد. دلیل دوم این اسـت که توزیع گاما دارای تعدادی ژن‌های عمده اثر با اثرات بزرگ‌تر نسبت به توزیع یکنواخت است. در نتیجه برآوردهای ارزش‌های اصلاحی ژنومی در صفاتی با تعدادی ژن‌های عمده اثر با اثرات بزرگتر، صحیح‌تر است.

یکی از ضعف‌های روش gblup این است که برای همه نشانگرها سهم یکسانی از واریانس را در پیش‌بینی ارزش‌های اصلاحی در نظر می‌گیرد. این در حالی اسـت که در روش‌های بیزی بر حسـب توزیع پیشین، وزن‌های متفاوتی به نشانگرها اختصاص داده می‌شود.

جدول ۵- صحت ارزش‌های اصلاحی ژنومی در تراکم نشانگری و تعداد qtl یکسان با توزیع واریانس یکنواخت و گاما

Table 5- The estimated genomic accuracy in same QTL numbers and SNP density with uniform and gamma variance distribution

توزیع واریانس Variance Distribution	وراثت پذیری Heritability	مدل آماری Statistical model GBLUP	BayesA	BayesB	BayesC	Bayes LASSO
یکنواخت Uniform	0.1	0.674(0.01)	0.677(0.01)	0.719(0.01)	0.715(0.02)	0.712(0.02)
	0.3	0.683(0.01)	0.685(0.01)	0.734(0.01)	0.730(0.02)	0.726(0.02)
	0.5	0.733(0.01)	0.740(0.01)	0.802(0.01)	0.801(0.02)	0.796(0.02)
گاما Gamma	0.1	0.684(0.01)	0.691(0.01)	0.760(0.01)	0.758(0.02)	0.753(0.02)
	0.3	0.695(0.01)	0.700(0.01)	0.779(0.01)	0.776(0.02)	0.776(0.02)
	0.5	0.739(0.01)	0.749(0.01)	0.845(0.01)	0.842(0.02)	0.841(0.02)

نتیجه‌گیری کلی

نتایج این تحقیق نشان داد که عواملی نظیر تراکم نشانگرها، تعداد qtl، توزیع واریانس QTL و وراثت‌پذیری صـفت بر میزان صـحت برآورد ارزش‌های اصـلاحی ژنومی مؤثر هسـتند. در صـفات با وراثت‌پذیری بالا هر چه تراکم نشـانگرها بیشـتر و تعداد qtl کمتر، میزان صحت برآورد ارزش‌های اصلاحی ژنومی نیز بیشتر است. در مطالعات ژنومی چنانچه صفتی تحت تأثیر تعداد کمی qtl قرار داشته باشـد، برآورد ارزش‌های اصـلاحی توسـط روش بیزb می‌تواند نتیجه مطلوب‌تری ارائه دهد. لیکن چنانچه صـفت تحت تأثیر تعداد زیادی qtl است، استفاده از هر پنج روش آماری gblup، بیزa، بیزb، بیزC و بیز lasso مزیت یکسـانی خواهد داشـت. حتی در بعضـی موارد روش gblup عملکرد بهتری نسـبت به چهار روش دیگر دارد. تراکم نشانگرها بر صحت ارزیابی ژنومی در صفات با وراثت‌پذیری متوسط تأثیری نداشت و از آنجا که بیشتر صفات اقتصادی در نژادهای بومی گوسفند از دسته صفات با وراثت‌پذیری متوسط هستند، می‌توان از ۵۰۰ تا ۱۰۰۰ نشـانگر برای برآورد ارزش‌های اصـلاحی ژنومی در برنامه‌های شبیه‌سازی استفاده نمود.

منابع

1- Buch, L. H., M. K. Sørensen, P. Berg, L. D. Pedersen, and A. C. Sørensen. 2012. Genomic selection strategies in dairy cattle: strong positive interaction between use of genotypic information and intensive use of young bulls. Journal of Animal Breeding and Genetics, 129: 138-151.
2- Calus, M. P. L., T. H. E, Meuwissen, A. P. W, De Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics, 178:553-561.
3- Chamberlain, A. J., H. C. McPartlan, and M. E. Goddard.2007. The Number of Loci That Affect Milk Production Traits in Dairy Cattle. Genetics, 177: 1117-1123.
4- Coster, A., J. W. M, Bastiaansen, M. P. L, Calus, J. A. M, Van Arendonk, and H. Bovenhuis. 2010. Sensitivity of methods for estimating breeding values using genetic markers to the num-ber of QTL and distribution of QTL variance. Genetic Selection Evolution, 42: 9-15.
5- Daetwyler, H. D., R. Pong-Wong, B. Villanueva and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185: 1021-1031.
6- Gilmour, A. R., R. Thompson, B. R. Cullis. 1995. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biomet, 51: 1440-1450.
7- Goddard, M.E. 2008. Genomic selection prediction of accuracy and maximization of long term response. Genetics, 136: 245-257.
8- Haldane, J. B. S. 1919. The combination of linkage values and the calculation of distance between the loci of linked factors. Genetics, 8: 299-309.
9- Meuwissen, T. H. E. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. Genetic Selection Evolution, 41: 35-47.
10- Meuwissen, T. H. E., B. Hayes, and M. E .Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819–1829.
11- Muir, W. M., G. K. S. Wong, Y. Zhang, J. Wang, M. A. M Groenen, R. Crooijmans, H. J. Megens, H. Zhang, R. Okimoto, A. Vereijken, A. Jungerius, G. A. A. Albers, C. T. Lawley, M. E .Delany, S. MacEachern, and H. H. Cheng. 2007. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. Proceedings of the National Academy of Sciences of the United States of America, 105: 17312-17317.
12- Nejati-Javaremi, A., C. Smith, and P. J. Gibson.1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. Journal of Animal Science, 75: 1738-1745.
13- R Core, T.2015. R A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
14- Saatchi, M., J. Ward, and D. J. Garrick. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. Journal of Animal Science, 91: 1538–1551.
15- Sargolzaei, M. and F.S. Schenkel.2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25: 680-681.
16- Solberg, T. R., A. K. Sonesson., J. A. Woolliams, and T. H. E. Meuwissen. 2008. Genomic selection using different marker types and densities. Journal of Animal Science, 86: 2447-2454.
17- Sved, J.A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theorical Population Biology, 2: 125-141.
18- Villumsen, T. M., L. Janss, and M. S. Lund.2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. Journal of Animal Breeding and Genetics, 126: 3-13.
19- Wimmer, V., C. Lehermeier, T. Albrecht, H. J .Auinger, Y. Wang, and C. C. Schön. 2013. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics, 195: 573-587.

Iranian Journal of Animal Science Research
Vol. 12, No. 2, Summer 2020, p. 241-250

نشریه پژوهشهای علوم دامی ایران
جلد ۱۲، شماره ۲، تابستان ۱۳۹۹، ص. ۲۵۰-۲۴۱

A Comparison of GBLUP and Bayesian Methods in Prediction of Genomic Breeding Values under Different Genetic Architectures

R. Seyed Sharifi[1*], F. Ala Noshahr[2], N. Hedayat evrigh[1], J. Seifdavati[1]
Submitted: 22-05-2019
Accepted: 29-09-2019

Introduction[1] Genomic Selection (GS) has been proved to be a powerful tool for estimating genetic values in livestock breeding. Newly developed sequencing technologies have dramatically reduced the cost of genotyping and significantly increased the scale of genotype data that used for GS. The estimation of breeding values in order to select the best animals as parents of the next generation is the main goal of animal breeding programs. Traditional methods of genetic evaluation were performed using a combination of phenotypic and pedigree information to produce estimated breeding values. Most simulation studies of genomic selection (GS) methods have considered genetic architectures in which the number and relative magnitudes of quantitative trait loci (QTL) have varied. Among the Bayesian methods, those using marker-specific shrinkage of effects (e.g., BayesA or BayesB of or the Bayesian LASSO are commonly used in animal breeding applications. The Bayesian methods proposed differ in the way of looking at the variances of parameters. In classical livestock breeding methods, selection for important economic traits using pedigree information with individual phenotypic records was performed and best Linear Prediction of Breeding Values (BLUP) is achieved. In genome selection, genomic breeding values of all individuals can be predicted with high accuracy using a linear model. Various factors can be affecting the accuracy of genomic breeding values. Therefore, the present study aimed to evaluate the accuracy of estimating genomic breeding values in different genetic architectures including different distributions of gene effects, different numbers of QTL, different levels of heritability and different marker densities using GBLUP and Bayesian methods including Bayes A, Bayes B, Bayes C and Bayes LASSO. In addition to comparing the performance of different methods in different genetic architectures, a marker density and QTL numbers were introduced for simulation programs of sheep populations.

Materials and Methods To create a basic population (G0), 100 heads of livestock, including 50 males and 50 females, were considered. The frequency of primary alleles for single-nucleotide polymorphisms in the basal generation was considered to be 0.5. To create the first generation (G1), the parents were randomly selected from the males and females of the G0 generation. Parental gametes were simulated based on the assumption of disconnection imbalance using the Halldan location function method, and then randomly generated gametes were randomly selected and mixed to create a new generation of G1 generation. A genome with a length of 300 cM was simulated and 500, 1000 and 1500 SNPs were equally spaced over the chromosome. Three different numbers of QTL (50, 100 and 150) were considered and QTLs were uniformly distributed over the chromosome. One hundred individuals, including 50 males and 50 females, were simulated for the base population. The first generation structure was followed through to the 50th generation of random mating to make linkage disequilibrium populations. Generation 51 was assumed as a training population and the other generations (52 to 60) as validation populations. Five methods, GBLUP, Bayes A, Bayes B, Bayes C and Bayesian LASSO, were used to estimate genomic breeding values.

Results and Discussion In all five methods, the accuracy of genomic values decreased as the number of QTLs increased from 50 to 150. The reason for this can be attributed to the limited amount of genetic variance distributed over many QTLs. Also predicting accuracy of all five methods increased with increasing marker density. Results showed that increasing marker density at low (0.1) and high (0.5) heritability levels, increased genomic accuracy but increasing at moderate heritability (0.3) traits did not affect the accuracy of genomic evaluation. Accuracy of genomic breeding values in the gamma distribution provides better gene effects to uniform distributions.

1-Associate Prof., Dept. of Animal Sciences, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran.
2-PhD graduated, Dept. of Animal Sciences, Faculty of Agricultural Sciences, University of Tabriz, Tabriz, Iran.
(*- Corresponding Author Email: reza_seyedsharifi@yahoo.com)
DOI:10.22067/ijasr.v12i2.80862

Conclusion The results showed that factors such as marker density, QTL numbers, distribution QTL effect and trait heritability were effective in estimating the accuracy of genomic breeding values. In high heritability traits, the higher markers density and lower QTL numbers, leading to increase accuracy of estimating genomic breeding values. In genomic studies, if the trait is affected by a small number of QTLs, estimation of breeding values by Bayes B method can yield a more favorable result. Marker densities did not affect the accuracy of genomic evaluation in traits of moderate heritability, and since most of the economic traits in native species of sheep are moderate heritability, 500 to 1000 markers can be used to estimate breeding values in simulation programs.

Keywords: BayesA, BayesB, BayesC, Bayesian LASSO, Genomic BLUP.